

INDEXAÇÃO AUTOMÁTICA POR INTELIGÊNCIA ARTIFICIAL: TRANSFORMAÇÕES NA REPRESENTAÇÃO DA INFORMAÇÃO DIGITAL

AUTOMATIC INDEXING BY ARTIFICIAL INTELLIGENCE: TRANSFORMATIONS IN THE REPRESENTATION OF DIGITAL INFORMATION

LAÍS CIRILO LIMA GARCEZ

Programa de Pós-Graduação em Ciência da Informação (PPGCI/UFG)

ORCID: <https://orcid.org/0009-0001-0057-806X>

ROR: <https://ror.org/0039d5757>

LAIS PEREIRA DE OLIVEIRA

Programa de Pós-Graduação em Ciência da Informação (PPGCI/UFG)

ORCID: <https://orcid.org/0000-0001-9092-4204>

ROR: <https://ror.org/0039d5757>

MODALIDADE: RESUMO EXPANDIDO

Resumo: Aborda a indexação de assunto sob a perspectiva da inteligência artificial em ambientes digitais. Objetiva compreender como a IA tem sido abordada na literatura científica da Ciência da Informação no que se refere à indexação. Trata-se de pesquisa descritiva qualitativa, baseada em revisão bibliográfica via Brapci e Google Scholar. Os resultados indicam que a IA é utilizada para ampliar a eficiência, a padronização e a escalabilidade da indexação, sem substituir integralmente a mediação humana. Conclui-se que há uma reconfiguração dos processos de representação da informação, marcada pela consolidação de modelos híbridos entre automação e análise humana.

Palavras-chave: inteligência artificial na indexação; informação digital – representação; ambientes digitais.

Abstract: This study addresses subject indexing from the perspective of artificial intelligence in digital environments. It aims to understand how AI has been approached in Information Science literature regarding indexing. This is a qualitative descriptive study based on a bibliographic review conducted in databases BRAPCI and Google Scholar. The results indicate that AI is used to enhance efficiency, standardization, and scalability in indexing processes, without fully replacing human mediation. It is concluded that there is a reconfiguration of information representation processes, marked by the consolidation of hybrid models combining automation and human analysis.

Keywords: artificial intelligence in indexing; digital information representation; digital environments.



1 INTRODUÇÃO

A evolução dos ambientes digitais intensificou os desafios relacionados à organização da informação (OI), especialmente diante do aumento da produção documental, fenômeno associado ao *Big Data*, entendido como um grande volume de dados passíveis de tratamento e uso. Nesse contexto, a indexação, historicamente baseada na mediação humana e no uso de vocabulários controlados, passa a incorporar sistemas de inteligência artificial (IA), capazes de automatizar e ampliar a escala do tratamento informacional por assunto.

A indexação configura-se como processo essencial nas unidades de informação, responsável por evidenciar o conteúdo temático dos documentos e influenciar diretamente os resultados da recuperação da informação (Fujita, 2003). A indexação automática, por sua vez, corresponde à substituição parcial da atividade humana por processos computacionais (Lancaster, 2004).

Observa-se uma transição entre a indexação manual, marcada pela subjetividade, e a indexação automática, caracterizada pela escalabilidade e eficiência. Contudo, essa mudança implica desafios, como a possível perda de profundidade semântica e a necessidade de lidar com o problema do *aboutness*, entendido como a identificação do assunto de um documento em seu contexto (Hjørland, 2001). Nesse cenário, evidencia-se uma reconfiguração dos paradigmas tradicionais de representação da informação diante do uso da IA.

Diante disso, coloca-se a seguinte questão: como a IA tem sido abordada na literatura científica da Ciência da Informação (CI) no que se refere à indexação? O presente estudo objetiva compreender essa relação a partir da literatura científica, buscando mapear abordagens, identificar tendências e discutir implicações para a organização e representação da informação em ambientes digitais.



2 METODOLOGIA

A pesquisa caracteriza-se como descritiva qualitativa, fundamentada em revisão bibliográfica sobre IA aplicada à indexação. A escolha da Brapci justifica-se por sua relevância para a produção científica em CI no contexto brasileiro, ao passo que o *Google Scholar* foi utilizado devido à sua ampla cobertura interdisciplinar e recuperação de literatura recente sobre tecnologias digitais.

A coleta de dados foi realizada entre março e abril de 2026, utilizando o termo fixo "inteligência artificial" combinado aos descritores "indexação automática" e "organização da informação", por meio das estratégias de busca "inteligência artificial AND indexação automática", "inteligência artificial AND organização da informação" e a combinação dos três termos simultaneamente. A pesquisa buscou trabalhos publicados, principalmente, entre 2020 e 2025.

Na Brapci, foram recuperadas 34 obras. Já no *Google Scholar*, a busca resultou em milhares de resultados, razão pela qual se adotou como critério a ordenação por relevância disponibilizada pelo próprio buscador. A seleção dos estudos seguiu critérios de inclusão (aderência temática, presença dos termos no título, resumo ou palavras-chave e disponibilidade do texto completo) e exclusão (duplicidade e ausência de relação direta com o tema). Considerando o caráter exploratório da pesquisa e os limites desta ocasião analítica, definiu-se uma amostragem por conveniência, restrita aos trabalhos com maior aderência temática e recorrência nos resultados recuperados. Os estudos selecionados foram organizados e analisados por meio de análise temática, visando identificar padrões, tendências e recorrências na literatura acerca da aplicação da IA na indexação.



3 RESULTADOS E DISCUSSÕES

A relação entre IA e indexação automática mostra-se recorrente na literatura recente da CI, especialmente a partir de 2021, movimento intensificado pela popularização dessas tecnologias após o lançamento do ChatGPT em 2022 (OpenAI, 2022). Nesse sentido, Nascimento, Martins e Albuquerque (2023) indicam que a automação é impulsionada pela necessidade de aumentar a eficiência, reduzir custos e mitigar a subjetividade inerente ao processo de indexação. A IA permite o processamento de um grande volume da informação, “[...] uma vez que se alcança um processo uniforme de diferentes documentos em curtos espaços de tempo” (p. 18). Destaca-se o Machine Learning (ML) como subcategoria promissora ao desenvolver o reconhecimento de padrões e aprendizagem contínua.

Contudo, essa incorporação não implica substituição plena do trabalho intelectual, mas como reconfiguração do processo. Indexação automática e semiautomática passam a ser compreendidas como formas de apoio ao indexador, combinando processamento computacional com validação humana, especialmente em contextos que exigem interpretação semântica mais refinada (Silva; Fujita, 2004; Guimarães, 2000).

Do ponto de vista técnico, retomando o texto de Nascimento, Martins e Albuquerque (2023), observa-se uma evolução dos métodos utilizados. Se inicialmente predominavam abordagens estatísticas baseadas em frequência de termos, a literatura recente aponta uma transição para métodos semânticos apoiados em IA e processamento de linguagem natural, capazes de lidar com estruturas linguísticas mais complexas.

Nascimento, Costa e Arakaki, (2025), na mesma linha, exploram sistemas de automatização documental a partir da utilização de *scripts* em Python a partir da associação de termos extraídos dos títulos documentados a conceitos do vocabulário controlado. A partir dessa experiência, é posto novamente o caráter eficaz da IA como redutor do esforço manual da indexação e o processamento de um volume informacional massivo com rapidez e precisão: “[...] A adoção de ferramentas baseadas em IA fortalece a infraestrutura da gestão da informação, tornando os processos de organização e recuperação da informação mais eficientes e beneficiando pesquisadores, bibliotecários e usuários finais” (p. 4).

Apesar dos avanços, a literatura também enfatiza limites importantes dessas tecnologias. Sistemas de IA ainda apresentam dificuldades na compreensão contextual e na interpretação conceitual profunda, sobretudo em domínios especializados, o que mantém a necessidade de intervenção humana qualificada (De Luca; Fallucchi; Morelato, 2021; Gomes, 2022). Além disso, problemas relacionados à qualidade dos



dados e ao processo de treinamento impactam diretamente o desempenho dos modelos, podendo comprometer a coerência dos descritores atribuídos (Géron, 2021). Consolida-se uma abordagem híbrida, em que a IA atua como elemento estruturante não autônomo, sendo a eficácia da representação dependente da articulação entre automação, controle terminológico e mediação humana (Pereira; Junior, 2025).

Por fim, Torres *et al.* (2025), colocam a IA **não** como substituta da análise humana – ela reconfigura e potencializa os métodos de indexação. A articulação entre diferentes especialistas — como cientistas da informação, historiadores, artistas, desenvolvedores de IA e estudiosos das artes visuais — revela-se essencial para a adaptação e o aprimoramento das heurísticas clássicas frente às exigências das tecnologias computacionais.



4 CONSIDERAÇÕES FINAIS

O objetivo da pesquisa foi compreender como a IA tem sido abordada na literatura científica da CI no que se refere à indexação. A partir da revisão realizada na Brapci e Google Scholar, observa-se que a IA é predominantemente tratada como um recurso de apoio à indexação, voltado à ampliação da eficiência, redução de custos e tratamento de grandes volumes de informação.

Os estudos por ora analisados dentro do plano amostral indicam que a incorporação da IA não configura uma substituição da indexação manual, mas uma reconfiguração do processo, marcada pela transição de métodos estatísticos para abordagens semânticas baseadas em aprendizado de máquina e processamento de linguagem natural. Nesse contexto, consolidam-se modelos híbridos, nos quais a automação e a mediação humana atuam de forma complementar.

Conclui-se que, embora a IA potencialize a indexação em termos de escala, padronização e velocidade, ainda apresenta limitações quanto à compreensão semântica profunda, à dependência de dados de treinamento e à transparência dos processos algorítmicos, o que mantém a centralidade do profissional da informação.

Além dos desafios operacionais, a literatura analisada evidencia limites epistemológicos relacionados à própria representação temática da informação. A substituição parcial da análise humana por modelos inferenciais baseados em padrões probabilísticos tensiona conceitos tradicionais da indexação, como o *aboutness*, uma vez que a atribuição temática passa a depender crescentemente de padrões estatísticos e não apenas de interpretação contextual. Assim, a IA não rompe integralmente com os fundamentos teóricos da organização do conhecimento, mas reconfigura seus modos de produção, mediação e representação, indicando a necessidade de aprofundamento crítico dos estudos sobre seus impactos na CI.



REFERÊNCIAS

DE LUCA, E.W.; FALLUCCHI, F.; MORELATO, R. Teaching an Algorithm How to Catalog a Book. **Computers**, v. 10(11), n. 155, 2021. Disponível em: <https://www.mdpi.com/2073-431X/10/11/155>. Acesso em 30 abr. 2026.

FUJITA, M.S.L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2089>. Acesso em: 30 abr. 2026.

GÉRON, A. **Mãos à obra**: aprendizado de máquina com Scikit-Learn, Keras e TensorFlow. Conceitos, ferramentas e técnicas para a construção de sistemas inteligentes. 2. ed. Rio de Janeiro: Alta Books, 2021.

GOMES, L. I. E. Transformação digital e Inteligência Artificial nos serviços de informação: inovação e perspectivas para a Ciência da Informação no mundo pós-pandemia. **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 15, n. 1, p. 148-166, 2022. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/41490>. Acesso em: 30 abr. 2026.

GUIMARÃES, J. A. C. **Indexação em um contexto de novas tecnologias**. [S.l.: s.n.], 2000. 10 p. Texto didático.

HJØRLAND, B. Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. **Journal of the American Society for Information Science and Technology**, Hoboken, v. 52, n. 9, p. 774-778, 2001. Disponível em: <https://browzine.com/libraries/134/articles/26187100>. Acesso em: 30 abr. 2026.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. rev. atual. Brasília: Briquet de. Lemos, 2004. 452 p.

NASCIMENTO, G. D. do; MARTINS, G. K.; ALBUQUERQUE, M. E. B. C. de. Automação da indexação: evidências e tendências da produção científica indexada na Brapci. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 28, p. 1-20, 2023. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/91956>. Acesso em: 30 abr. 2026.

NASCIMENTO, R. L. S.; COSTA, B. C. C.; ARAKAKI, A. C. S. Aprimoramento do Annif para indexação automática de publicações em português. *In*: WORKSHOP DE INFORMAÇÃO DADOS E TECNOLOGIA, 8., Marília, 2025. **Anais** [...]. Marília: UNESP,



2025. Disponível em: <https://widat.ibict.br/index.php/widat2025/article/view/273>. Acesso em: 30 abr. 2026.

OPENAI. **ChatGPT**: optimizing language models for dialogue. 30 nov. 2022.

PEREIRA, J. C. S. S. J.; JUNIOR, R. F. G. Inteligência artificial aplicada à organização do conhecimento: uma taxonomia. *In*: ISKO BRASIL, 8., Canela, 2025. **Anais** [...]. Canela, RS: ISKO Brasil, 2025. Disponível em: <https://isko.org.br/ojs/index.php/iskobrasil/article/view/112>. Acesso em: 30 abr. 2026.

SILVA, M. R.; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, Campinas, v. 16, n. 2, p. 133-161, maio/ago., 2004. Disponível em: <https://periodicos.puc-campinas.edu.br/transinfo/article/view/6373>. Acesso em: 30 abr. 2026.

TORRES, A. A. L.; MACULAN, B. C. M. S.; ROCHA, A. A.; ASSUNÇÃO, F. M.; MARQUES, F. B.; SILVA, G. R. Inteligência artificial e a indexação de imagens com o Método Iconográfico de Panofsky. *In*: ISKO Brasil, 8., Canela, 2025. **Anais** [...]. Canela, RS: ISKO Brasil, 2025. Disponível em: <https://isko.org.br/ojs/index.php/iskobrasil/article/view/116>. Acesso em: 30 abr. 2026.